AD-A206 740

DTIC ELECTE APR 12 1989 S H

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION	1b. RESTRICTIVE MARKINGS
	DTIC FILE COPY
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE NA	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NA	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION: University of British Columbia	6b. OFFICE SYMBOL (If applicable):
6c. ADDRESS (City, State, and ZIP Code): Dept. of Biochemistry, 2146 Health Sciences Mall, Vancouver, B.C., Canada V6T 1W5	7a. NAME OF MONITORING ORGANIZATION
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: Office of Naval Research	8b. OFFICE SYMBOL (If applicable): ONR
8c. ADDRESS (City, State, and ZIP Code): 800 N. Quincy St., Arlington, VA 22217-5000	7b. ADDRESS (City, State, and ZIP Code)
9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: N00014-87-G-0049	
10. SOURCE OF FUNDING NUMBERS	

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
61153N	RR04106		

11. TITLE (Include Security Classification): Molecular Biology of Archaebacteria

12. PERSONAL AUTHOR(S): Patrick P. Dennis

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Annual	FROM 4/88 TO 3/89	890331	

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Transcripts derived from the 5.2 Kbp fragment of Halobacterium cutirubrum genomic DNA that encodes the L11e, L1e, L10e and L12e ribosomal protein genes have been characterized. The leader region of the L1e, L10e, 112e tricistronic transcript contains a region that exhibits sequence and stuctural similarity to/ the L1e binding site in 23S rRNA. This leader sequence is presumably used to autogenously regulate mRNA translation. The gene encoding the H. cutirubrum superoxide dismutase has been cloned and sequences and its transcript has been characterized. Transcription of the gene is inducible by oxygen radicals. The promoter regions from the two rRNA operons of H. Marismortui have been sequenced and the promoters have been defined by nuclease protection. The 5' flanking region of these two operons differ substantially with regard to (i) the number of promoters, (ii) the processing signals, and (iii) nucleotide sequence of tne first 130 nucleotide within the 16S gene. Finally, a family of five genes encoding multiforms of the eucaryotic L12e protein and a single L10e protein have been cloned, transcript mapped, and compared to the corresponding eubacterial and archaebacterial proteins.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ OTIC USERS	(AW)
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code) / 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete. SECURITY CLASSIFICATION OF THIS PAGE

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

Abbreviations: Hcu - Halobacterium cutirubrum
Hma - Halobacterium marismortui
Sso - Sulfolobus solfataricus
Eco - Escherichia coli
Sce - Saccharomyces cerevisiae

Research Objectives

(i) to characterize the principles of gene organization and regulation of gene expression in archaebacteria; (ii) to elucidate the evolutionary relationship between these novel organisms and the traditional eubacterial and eucaryotic organisms; (iii) to understand in biophysical and molecular terms some of the mechanisms that allow archaebacteria to inhabit extreme environments.

There is currently considerable controversy concerning the evolutionary origins of archaebacteria and their relationship to eubacteria and eucaryotes (1,2). Most investigators believe that archaebacteria are a monophyletic group, separate and distinct from eubacteria and eucaryotes. However, Lake (1) has suggested that they are polyphyletic; he believes that the halophiles are related to the eubacteria and that the sulfur-dependent thermoacidophiles are related to eucaryotes. Others believe this result to be artifactual (2). To address this problem we have been studying a group of four ribosomal proteins (and the genes that encode them) that have a well defined and highly conserved structure and function on the ribosome.

Ribosomal protein genes: The ribosoml A protein complex forms the stalk structure on the large ribosome subunit and is comprised of four copies (two dimers) of L12e and one copy of L10e ribosomal protein. The L10e protein binds to ribosomal RNA either directly or through the L11e protein which forms a bulge at the base of the stalk. This domain on the large subunit is the site of factor binding and associated GTPase activities during the protein synthesis cycle and is a conserved and defined feature on the ribosome from all organisms (3). A fourth protein, L1e, forms a distinct ridge on the large subunit near the peptidyl transferase centers and is involved directly in the interaction between the peptidyl tRNA and the P site and indirectly with the GTPase center. In Eco the genes encoding the L11, L1, L10 and L12 ribosomal proteins are located within a 3.0 Kb region of genomic DNA. This region has been cloned, sequenced and extensively characterized (4,5).

Using synthetic oligonucleotides based on known protein sequence, the L11e, L1e, L10e and L12e genes from two divergent archaebacterial species, Hcu and Sso, were cloned on 5.2 Kbp ClaI-BamHl and 6.2 Kbp EcoRI-BamHl genomic restriction fragments respectively. The ribosomal protein genes and other potential coding regions were revealed by sequence analysis. Remarkably for the two archaebacterial species the clustered arrangement of the L11e, L1e, L10e and L12e genes was identical to the arrangement in the eubacterium E. coli. The transcriptional mechanisms for controling the expression of the gene in the three organisms are different. Likewise, the coding sequences 5' and 3' to the tetragenic cluster are also totally unrelated between the three organisms.

Progress - Year 2

The in vivo transcripts derived from the H. cutirubrum 5.2 Kbp Cla-BamHl genomic restriction fragment have been characterized by oligonucleotide hybridization, northern hybridization, nuclease protection and primer extension analysis. These results are summarized and compared to E. coli transcripts in the accompanying figure 1. Briefly, the major L11e transcript is monocistronic and initiates at the A of the ATG translation initiation codon. A ten-fold less prominent NAB-L11e bicistronic transcript is also detected; this transcript is initiated at a G residue immediately in front of the ATG translation initiation codon. The intergenic space between NAB and L11e is 73 nucleotide long. Transcripts exiting the L11e gene terminate efficiently at a number of sites in the 203 nucleotide long L11e-L1e intergenic space. This intergenic space also contains the promoter responsible for production of the abundant L1e-L10e-L12e tricistronic mRNA. The 5' untranslated leader of this transcript is 74 nucleotides in length, is endowed with inverted repeat symmetry and exhibits sequence and structural similarity to the L1e binding site in 23S rRNA (Figure 2). The L1e-L10e and L10e-L12e intergenic space are only one and five nucleotides in length respectively. We have suggested that the 5' leader sequence on the Hcu L1e-L10e-L12e tricistronic rRNA might be utilized as a site for autogenous translational regulation by the L1e protein. In the eubacterium E. coli the L1 protein autogenously regulates translation of the L11-L1 mRNA and L10 regulates translation of the L10-L12 mRNA. When there is a deficiency of rRNA, the free L1e that accumulates can bind to the leader region of its own rRNA and prevent further translation. This insures a balance between the production of rRNA and L1e, L10e and L12e; how L11e synthesis is regulated in halobacteria remains to be determined.

The promoters for these ribosomal protein operons retain some of the same conserved sequences that are found in the tandem rRNA promoters. The conserved blocks are TTCGA and TTAA and are centered about 40 and 30 nucleotides in front of the transcription start sites. Termination occurs at T tracts in the (+) strand of the DNA; the efficiency of termination increases with the length of the tract. In the r-protein coding region tracts of multiple T residues on the (+) strand occur much less frequently than expected. Antitermination is important in polycistronic transcripts where balanced synthesis of the protein products is important.

Transcripts have not yet been characterized in S. solfataricus. It is, however, interesting to note that the intergenic spaces between the L11e, L1e, L10e and L12e genes are -1, -1 and 44 nucleotides respectively and that these are closely associated upstream and downstream open reading frames. This suggests that the four ribosomal protein genes and the adjacent open reading frames may be part of a large transcription unit.

Alignments between the deduced amino acid sequences of these archaebacterial ribosomal proteins to the available homologous protein sequences from eubacteria have been made. The alignments suggest that archaebacteria are a coherent phylogenetic group and that the Hcu and Sso proteins, L11e, L1e and L10e are approximately equal in evolutionary distance from the corresponding Eco proteins. Alignment of L12e protein is much more complex. The archaebacterial and eucaryotic proteins align end to end whereas

the eubacterial protein has clearly undergone major domain rearrangements during the early stages of eubacterial evolution. But again within the conserved globular domain the Hcu and Sso L12e proteins are more closely related to each other than either is to eubacterial or eucaryotic proteins.

Information on eucaryotic L10e and L12e genes and the proteins they encode is essential in providing a complete evolutionary perspective on the A protein complex. With this in mind we prepared an oligonucleotide and cloned five hybridyzing fragments of Saccharomyces cerevisiae genomic DNA. Four of the fragments encode a family of related L12e protein and the fifth encodes a L10e protein. Three of the L12e and the L10e genes are uninterrupted whereas the fourth L12e gene contains a 301 bp intron between codons 38 and 39 (alignment position 45-46). The four L12e proteins divide into two classes (I and II). The class I proteins (A and B) lack tryptophan and usually contain arginine at alignment position 46. The class II proteins (A and B) have an extended N-terminus, contain tryptophan at position 47 and lack arginine. Other eucaryotes apparently have only two L12e genes, one class I type and one class II type. The single archaebacterial L12e protein is more closely related to the eucaryotes L12eI. This suggests that there has been a recent duplication of both the type I and II genes in the yeast lineage. Comparison indicates that the original duplication to produce type I and II genes is an ancient event and probably occurred in the common primordial ancestor. This means that one of the genes (probably the type II gene) has been lost in the eubacterial and archaebacterial lineage (both the eubacterial and the archaebacterial L12e proteins contain the conserved arginine and lack tryptophan). In the eubacteria the remaining L12e gene has undergone substantial rearrangements.

We have also shown that in eucaryotes and in archaebacteria, a three quarter copy of an L12e sequence is fused to the carboxy terminus of L10e. In Eco this homology between L10 and L12 does not exist because during eubacterial evolution the L10 has been prematurely truncated at its C terminus and the L12 has been rearranged.

rRNA operons: In contast to H. cutirubrum, H. marismortui contains two (rather than one) rRNA transcription units. These units have been independently cloned on a 10 Kbp HindIII fragment and an 8 Kbp HindIII-ClaI fragment. The promoter regions of the two operons were sequenced and transcription initiation sites were characterized by nuclease protection analysis. The HC8 operon has three tandem promoters whereas the HH10 operon has a single promoter. In addition, the HC8 operon contains a typical archaebacteria processing signal in the 5' leader sequence whereas the HH10 operon appears to lack this processing signal. The proximal 130 nucleotides of each 16S gene has been sequenced. Surprisingly, there are many differences between the two genes. Normally within an organism multiple rRNA genes encode identical or nearly identical rRNA products. The characterization of these two operons is continuing.

Superoxide dismutase: Ribosome component genes are essential. As a contrast, we have been characterizing the non-essential gene encoding superoxide dismutase. The SOD enzyme was initially purified from H. cutirubrum and an amino terminal sequence was determined. The gene encoding SOD (sod) was cloned, sequenced and shown to encode a 200 amino acid long polypeptide that is related to the Fe and Mn SODs of eubacteria and unrelated to the CuZn enzyme of eucaryotes. The transcript of the sod gene has been characterized by northern hybridization, nuclease protection and primer

extension. The transcript is mono-cistronic; it initiates 2-3 nucleotides in front of the ATG translation initiation codon and terminates in a T tract about 40 nucleotide beyond the coding region. The 5' flanking region lacks easily recognizable elements that are normally present in archaebacterial promoters. The activity of the sod gene has been shown to be regulated; addition of paraquat, a generator of oxygen radicals, causes an increase in SOD activity and sod mRNA.

When a restriction fragment encoding the sod gene was hybridized to genomic DNA, it hybridized to the authentic gene and to a second related sequence. The related fragment has been cloned and partially characterized. This sequence appears to contain an open reading frame designated slg (SOD related gene sequence). The function of slg and the potential protein it encodes needs to be determined.

References

1. Lake, J. (1988) Nature 331, 184-186.
2. Olsen, G. and C. Woese (1989) Canad. J. Microbiol. 35, 119-123.
3. Lipjas, A. (1982) Prog. Biophys. Mol. Biol. 40, 161-228.
4. Post et al. (1979) PNAS 76, 1697-1701.
5. Downing, W. and P.P. Dennis (1987) J. Mol. Biol. 194, 609-620.

Objectives for Year 3

Ribosomal protein genes: We would like to begin to characterize the putative L1e binding site in the leader region of the L1e, L10e, L12e mRNA to determine if it can function in translation regulation. We will try to reconstruct the regulatory system in E. coli where genetic and physiological manipulation is more convenient.

Superoxide dismutase: We will complete the characterization of the slg clone in H. cutirubrum and determine its precise relationship to sod. We also intend to characterize the sod gene from H. volcanii and begin to isolate regulatory mutants affecting the activity of this gene.

rRNA operons: We are continuing to characterize the expression of the Sso rRNA operons. We are also continuing to sequence the two 16S genes from Hma, to study their transcription and their processing.

Publications

1. Shimmin, L.C. and P.P. Dennis (1989) Characterization of the L11, L1, L10 and L12 equivalent ribosomal protein gene clusters of the halophilic archaebacteria Halobacterium cutirubrum. EMBO Journal, in press.

2. Shimmin, L.C., C. Ramirez, A. Matheson and P.P. Dennis (1989) Sequence alignment and evolutionay comparison of the L10e and L12e ribosomal protein from archaebacteria, eubacteria and eucaryotes. J. Mol. Evol., in press.

3. Shimmin, L.C., C.H. Newton, C. Rosemary, J. Yee, W.L. Downing, A. Louise, A. Matheson and P.P. Dennis (1989) Organization of gene encoding L11, L1, L10 and L12 equivalent ribosomal protein in eubacteria, archaebacteria and eucaryotics. Canad. J. Microbiol. 35, 164-170.

4. Ramirez, C., L.C. Shimmin, C.H. Newton, A. Matheson and P.P. Dennis (1989) Structure and evolution of the L11, L1, L10 and L12 equivalent ribosomal protein in eubacteria, archaebacteria and eucaryotics. Canad. J. Microbiol. 35, 234-244.

5. May, B., P. Tam and P.P. Dennis (1989) The expression of the superoxide dismutase gene in Halobacterium cutirubrum and Halobacterium volcanii. Canad. J. Microbiol. 35, 171-175.

6. Mevarech, M., S. Hirsch-Twizer, S. Goldman, E. Yakobson, H. Eisenberg and P.P. Dennis (1989) Isolation and characterization of the rRNA gene clusters of Halobacterium marismortui. J. Bacteriol., in press.

7. C.H. Newton, J. Yee, L.C. Shimmin and P.P. Dennis (1989) A family of genes encode the multiple forms of the yeast L12 equivalent ribosomal proteins and a single form of the l10 equivalent ribosomal protein. Submitted for publication.

8. May, B. and P.P. Dennis (1989) Evolution and regulation of the gene encoding superoxide dismutase from the archaebacteria Halobacterium cutirubrum. J. Biol. Chem., in press.

Training

Bruce May: MRC, predoctoral fellow, Canadian
Lawrence Shimmin: ONR, predoctoral student, Canadian
Craig Newton: ONR technician, Canadian
Diedre de Jong-Wong: ONR technician, Canadian
Phelgun Joshi: ONR predoctoral student (grant), Canadian
Peter Durovic: MRC predoctoral student (grant), Canadian
Janet Yee: NSERC predoctorial student, Canadian

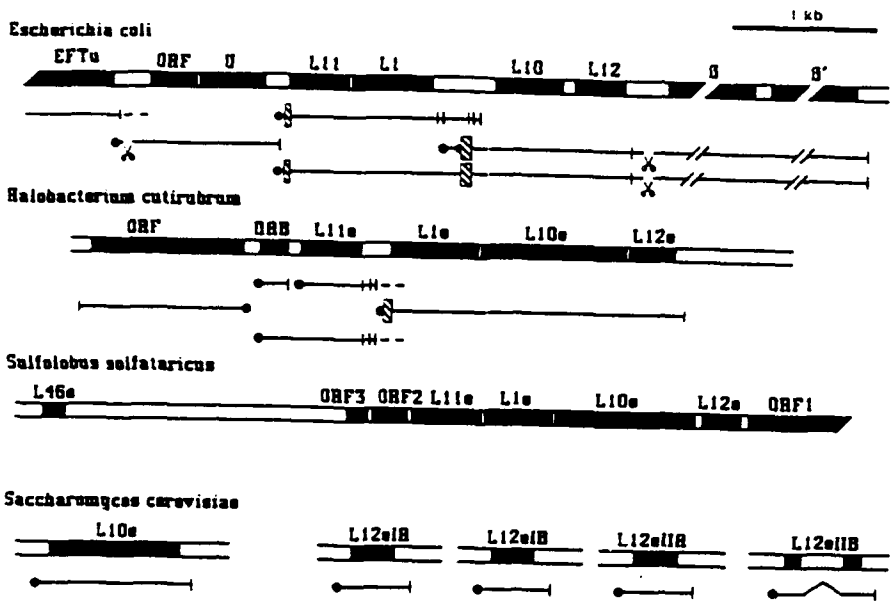

FIG. 1. The organization of genes encoding components of the ribosomal A protein complex in eubacteria, archaebacteria, and eucaryotes. Genomic clones containing the L10e and L12e genes from the eubacterium *E. coli* (top), the archaebacteria *H. cutirubrum* and *S. solfataricus* (middle), and the eucaryote *S. cerevisiae* (bottom) have been sequenced and characterized. Genes are indicated as solid boxes; unshaded regions are intergenic noncoding sequences. Other genes or open reading frames (ORF) in the neighbourhood of the L10e and L12e genes are indicated. The *in vivo* transcripts derived from the *E. coli*, *H. cutirubrum*, and *S. cerevisiae* gene have been characterized by oligonucleotide hybridization, nuclease protection, and primer extension analysis and are indicated below the physical maps. Major 5' transcript ends (●) and 3' transcript ends (|) are indicated. Stippled lines beyond major 3' transcript ends indicate a small but detectable amount of readthrough transcription. The scissored interruptions in the *E. coli* transcripts are RNase III dependent processing sites. In the *S. solfataricus* clone, the sequence between L46e and ORF 3 is incomplete and transcripts have yet to be characterized; all the coding regions on this fragment are oriented left to right. The *S. cerevisiae* L12eIIB gene is interrupted by a 301-bp intron between codons 38 and 39.